Check 2018, Alfredo.pdf

by Arta Sundjaja

Submission date: 20-Aug-2019 10:39AM (UTC+0700)
Submission ID: 1102514169
File name: 2018,_Alfredo.pdf (762.19K)
Word count: 2751
Character count: 14947

Data Warehouse Development For Flight Reservation System

Yoel Frans Alfredo
Computer Science Department, BINUS Graduate Program-Master of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
yoel.frans@binus.ac.id

Abba Suganda Girsang
Computer Science Department, BINUS Graduate Program-Master of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
agirsang@binus.edu

Sani Muhamad Isa
Computer Science Department, BINUS Graduate Program-Master of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
sani.m.isa@binus.ac.id

Ahmad Nurul Fajar
Information Systems Management Department, BINUS Graduate Program-Master of Information Systems Management
Bina Nusantara University
Jakarta, Indonesia 11480
afajar@binus.edu

***Abstract*—XYZ company is an online travel company that provides services for booking hotel and flight. As the companies grow fast, they need fast report to help the management create a decision for developing their product. They need this report to help them decide what the action they need to do based on the data they have for compete with fellow rival in online travel company. This data warehouse development has purpose to help provide a big picture of XYZ company's customers behavior through data in report generated. The data warehouse development uses some steps designed by Kimball. Information generated will be presented in dashboard or report. The data warehouse is expected to be able to provide quick and accurate information for management to choose best decision on next product development based on customer behavior.**

Keywords— data warehouse, reservation system, travel, flight, reservation

I. INTRODUCTION

XYZ is an online travel company which has focus on flight and hotel reservation services. Right now, XYZ company growth is fast and they ready to compete with their fellow rivals. The management need a report to support their decision through data analytics. Currently, their transaction data only stored in database and cannot be used by management on decision-making for their product development. Management usually do request to their development team directly if they need specific report request. Usually it might take lot time to process the report that requested by management from current transactional database.

Data warehouse is a tool which store all operational data and can be used to get insight from every information in there [1] [2] . Data warehouse aims to provide some dimensions analysis from several resources [3] [4] [5] . Usually data warehouse will store collection of historical and current data. Data from sources should be transformed into one consistent format and stored in data warehouse. The transformation will use ETL (Extract, Transform & Load) process to transform all data from sources into same format file before inserted into a warehouse [6] [7] [8]. The methodology that will be used for developing data warehouse is Kimball Lifecycle methodology which was conceived since mid of 1980 . This methodology is still used in a lot fields of research until now [9] [10].

Lot of research in various fields using data warehouse as tools to solve data processing problems. Tagirova uses data warehouse to improve storage subsystem as a part of the intellectual oil control system [11]. The data warehouse reduces labor costs for the development and provide the ability to integrate with existing high level information system. Tamer uses data warehouse to manage tourist data in Egypt [12]. The data warehouse integrate all available data sources and develop information based on reports generated to help decision makers in the tourism sector of Egypt make a decision using different point of views. Imen [13] uses data warehouses to help company's decision maker to identify information about their company and customer using opinion analysis through social media.. The model developed will predict the success of products prospects from the company itself. Kraus uses data warehouse to help detect a fraud in supply chain [14]. The design of data warehouse will use Benford's Law in order to detect fraud with the help of parameterized stored procedures.

II. THEORITICAL BACKGORUND

A. Data Warehouse Architecture

Data warehouse architecture have several tiers. The back-end tier, data warehouse tier, OLAP tier and front-

978-1-5386-9422-0/18/$31.00 ©2018 IEEE

end tier [3]. The back-end tier is the tools of extraction, transformation and loading (ETL) that feed data from internal or external sources into data warehouse. All data will be transformed into a unified format for data warehouse. The data warehouse tier is an enterprise data warehouse and/or collection of data marts and metadata repository that store all data transformed through ETL process. The OLAP tier is a tool which provides multidimensional view of the data from data warehouse. The front-end tier is tool for data analysis and visualization of the data through report or dashboard.

B. *Online Analytical Processing (OLAP)*

OLAP is a multidimensional view to analyze informations on different business perspective (dimensions) [4]. It consist of a creation of cubes to explore the informations in data warehouse or in the datamarts. Cube in OLAP is a multidimensional data structure (actual / virtual) that allows fast data analysis. The data arrangement on the cube serves to overcome the limitations of relational databases. Relational databases are not suitable for quick analysis of large amounts of data. OLAP functionality is to improve end-user productivity in business areas, such as IT developers and entire organizations. More oversight and timely access to strategic information can result in more effective decision-making.

Star schema is one among others data warehouse schema. This schema using fact table and dimension table to describe business requirements. The fact table is located in center and connected to all the dimension table that surrounds it. Dimension tables have attributes that help describing the dimensional value. number of attributes. The size of dimension tables are generally smaller than the fact table. The fact table should contains foreign keys that refer to dimension table, and attribute measure that can be aggregate according to analysis requirement. Star schema have several characteristics such as simple structure, less join rather than transactional schema, denormalized data and supported by business intelligence tools.

III. METHODOLOGY

A. *Preparing Data Source*

Data source is transaction data from transactional database that consist all data which related to flight reservations. Data which related to flight reservations is collected and stored in a new transactional database to make it easier to update the data regularly since the data is stored in same database with hotel reservations data. The data itself will be transformed through ETL process into a unified data structure.

B. *Design of Data Warehouse*

Data warehouse design will use Kimball methodology method that have 4 steps : choose the business process, declare the grain, identify dimensions and identify the fact [15].

- **Choose the process**. At this stage the process is to determine the business process and measurement used in the data warehouse. The business process of XYZ company that chosen in flight reservation transaction.
- **Declare the grain**. Once the business process has been identified, development and design team should pick the granularity of fact table. It's important to define what a fact table row is in the proposed business process dimensional model. XYZ company will use reservation as the lowest granularity in the fact table.
- **Identify dimension**. Once the grain of the fact table is established, the next step is choose the dimensions. Choose which dimensions that will be used to create a table fact according to the relationship of each row of data. The grain itself will often determine a primary or minimal set of dimensions. Dimensional table of flight reservation transactions are: flight dimensions, departure dimensions, destination dimensions, time dimension, payment dimension and payment gateway dimension.
- **Identify the fact**. On this stage, fact tables determined based on dimension and granularity that have been choosen before. Measure should be defined on this step, it will be stored in fact table. XYZ company uses some measurement in the fact table defined : total passenger count, total segment count, total male passenger, total female passenger and total amount/sales. The information than can generated from datawarehouse can be seen in this measurements. Proposed data warehouse uses transactional database with duration of three years from 2015 to 2017 and be extracted and transformed using the ETL process.

C. *Star Schema Model*

This model is describing business process choosen represented by the dimensions tables and fact table. Fact table will be the central of this model and dimension tables are connected to the fact table. Pictures of the star schema of the flight reservation process can be seen in Fig. 1. In the picture can be seen star schema consist of 6 dimension tables and 1 fact table. Dimension table consists of : flight dimensions, departure dimensions, destination dimensions, time dimension, payment dimension and payment gateway dimension and fact table consist of table fact reservation.

D. *Extract Transform Load*

This step is integration process from data sources into data warehouse. In the integrations process there's step that also known as extract, transform and load. Data will be extracted from the sources and transform into unified format that will be used in data warehouse. Once data transformed successfully, it will loaded into data warehouse.

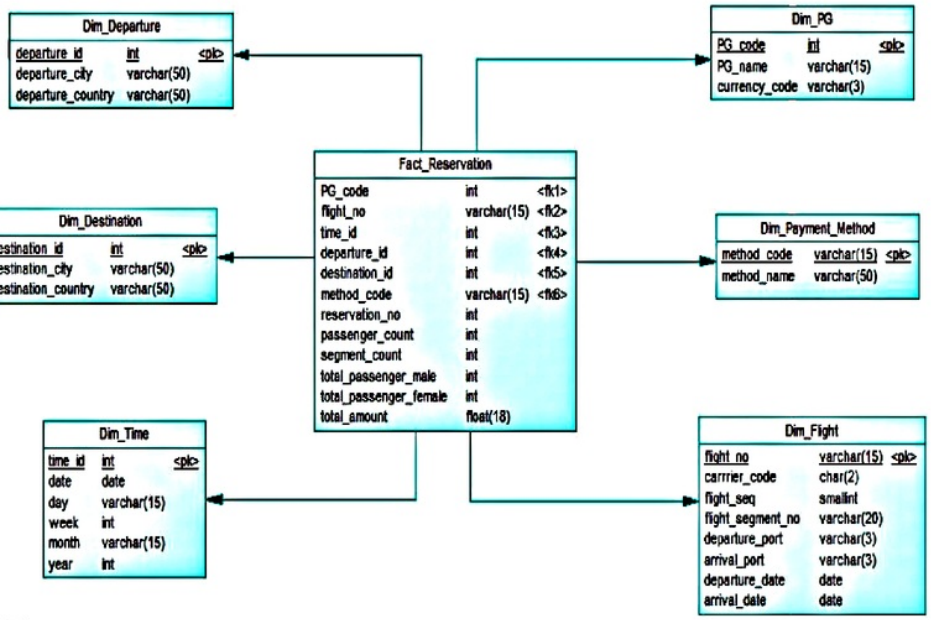


Fig. 1. Star Schema

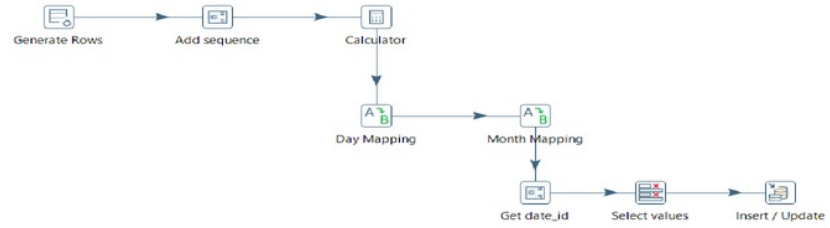


Fig. 2. Load Time Dimension Process

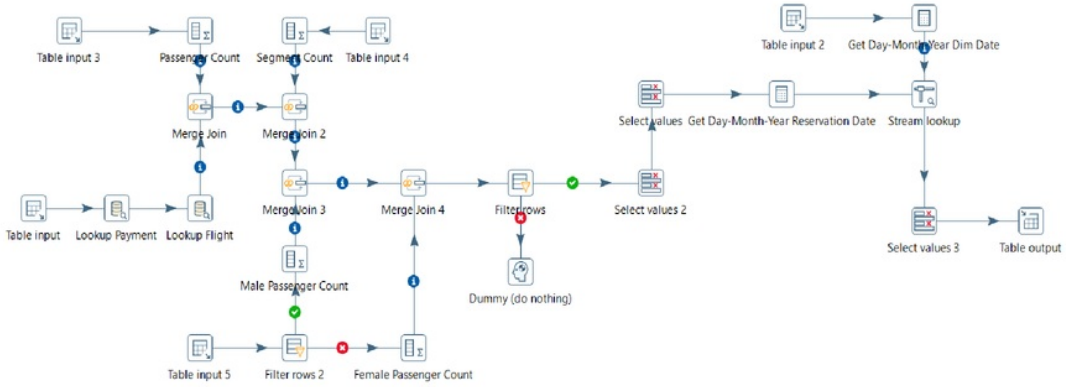


Fig. 3. Load Fact Reservation

E. *Designing OLAP Cube*

After the ETL process is successfully executed, OLAP cube should be designed. The cube should be designed to build OLAP analysis. In the cube hierarchy and measure should be defined.

IV. Analysis Result

The Business Intelligence development start from the star schema design that shown in Figure 1. Through ETL process, data from chosen star schema will be processed from Online Transactional Processing (OLTP) into Online Analytical Processing (OLAP).

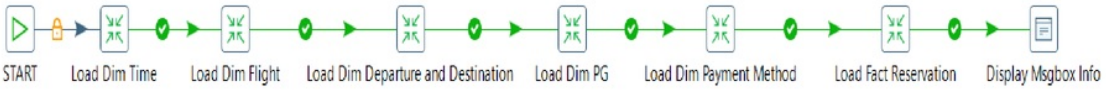
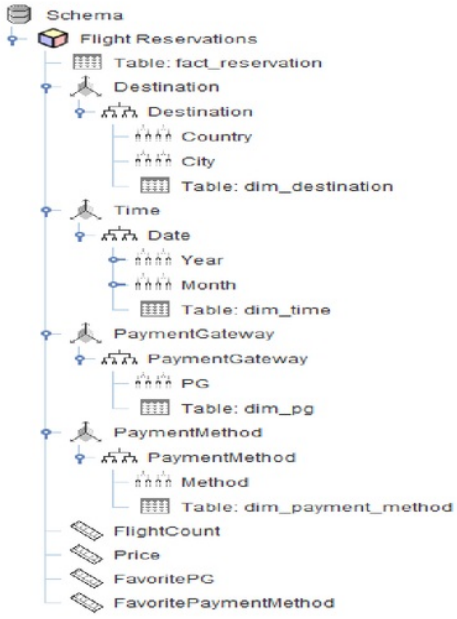


Fig. 4. ETL Job



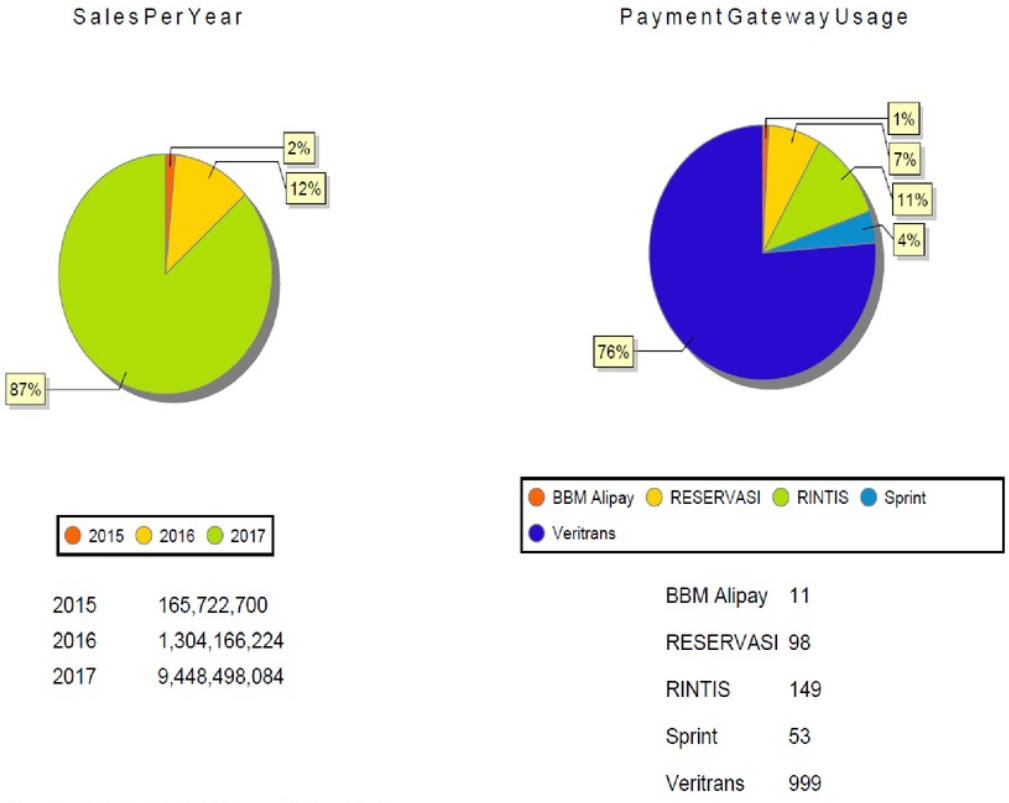
As we can see from Fig. 2, we generate date dimension process since. It generated from date attribute from OLTP data and split it value into day, date, month, and year then load it into time dimension table. Each dimension will be processed and loaded into created table. Once dimensions table already processed and the staging table is ready then the ETL process of fact table can be started. As we can see in Fig. 3.

To simplify the process, we can build ETL job that combine all process into one. Once we start the job, all process will be started and data will be generated and ready to be used. This job process flow can be seen in Fig. 4. Before we able to do OLAP analysis, OLAP cube should developed. In OLAP cube we can set hierarchy from available dimension and measure that will be used based on the requirement from PT. XYZ. Design of OLAP cube can be seen in Fig. 5.

Fig. 5. OLAP Cube Design

(All)	All Time.Dates					
Year	2015		2016		2017	
Country	FlightCount	Price	FlightCount	Price	FlightCount	Price
Australia	-	-	-	-	6	146,167,300
Bangladesh	-	-	-	-	2	27,303,300
Cambodia	-	-	-	-	5	95,416,580
China	-	-	-	-	4	151,062,000
Hong Kong	-	-	7	64,585,800	29	466,937,537
India	-	-	-	-	8	144,523,600
Indonesia	27	165,722,700	234	1,117,951,240	652	3,706,548,723
Japan	-	-	2	40,838,000	16	264,170,700
Korea, Republic Of	-	-	1	3,998,000	-	-
Macau	-	-	2	16,620,284	7	199,503,596
Malaysia	-	-	7	27,220,900	93	947,887,689
New Zealand	-	-	-	-	1	23,979,000
Philippines	-	-	-	-	17	720,719,908
Singapore	-	-	5	22,874,500	96	649,788,313
Spain	-	-	-	-	1	1,597,500
Taiwan	-	-	-	-	10	215,358,800
Thailand	-	-	3	10,077,500	61	1,480,990,827
Timor Leste	-	-	-	-	5	44,158,100
United States	-	-	-	-	1	37,370,100
Vietnam	-	-	-	-	6	117,017,011

Fig. 6. OLAP Analysis by Destination and Years



2015	165,722,700
2016	1,304,166,224
2017	9,448,498,084

BBM Alipay	11
RESERVASI	98
RINTIS	149
Sprint	53
Veritrans	999

Fig. 7. Report on Sales per Year and Payment Gateway Performance

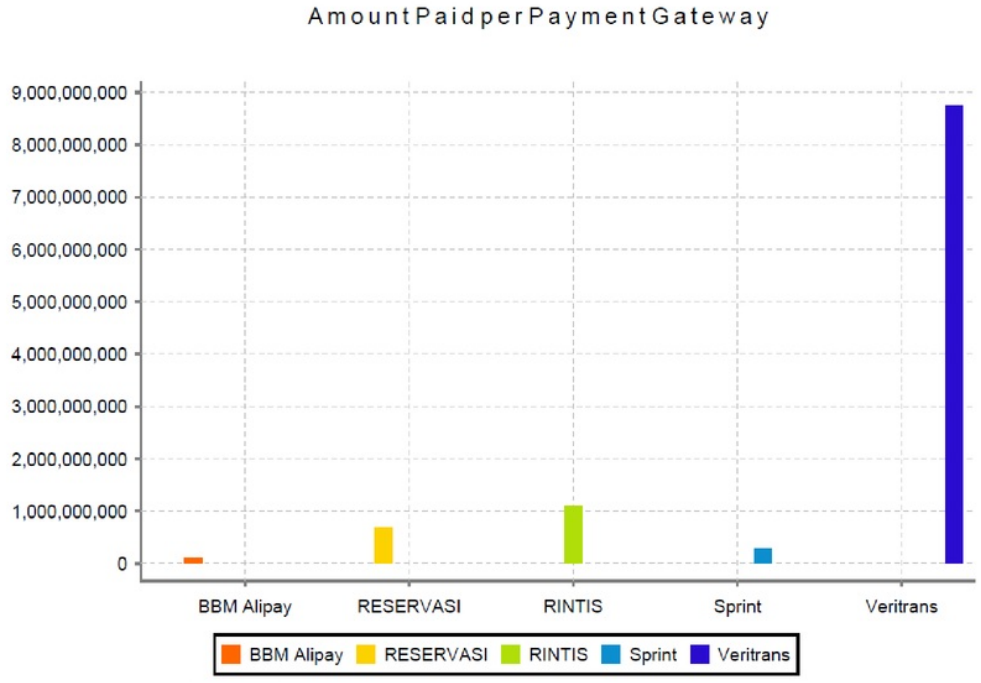


Fig. 8. Report on Gross Merchandise Volume from Payment Gateway

use OLAP analysis. One example, we can see a drill-down analysis on flight booked and total Gross Merchandise Volume based on country destination and years. We can see the example of drill-down analysis on Figure 6. From the report, we can see how much total Gross Merchandise Volume we have each year and how the performance of available payment gateway on PT. XYZ.Top management can make a decision to keep or change the payment gateway based on their performance and tendency of user. Besides, top management can see the performance of each team on the company based on the Gross Merchandise Volume, do they achieve annual target or not and make a decision on that.

On the current system, top management can not see live report of the company performance. They still need to ask operation and finance team to prepare and calculate how much Gross Merchandise Volume they have achieved which can be a miscalculation due to human error. Using developed OLAP Analysis, dashboard and report can be designed. On of the report that can be generated from OLAP analysis is sales (Gross Merchandise Volume) report per year, payment gateway performance, and total Gross Merchandise Volume per payment gateway. We can see the report example on Figure 7 and Figure 8. Besides the report, through OLAP analysis we can build a dashboard to help team and top management to see the performance of the company on real-time. This dashboard consist of three charts : sales per country, sales per year, and flight count per destination country as we can see on Figure 9.

V. Conclusion

With data warehouse developed, XYZ company can make a better decision through analysis on the insight from information that presented from OLAP analysis. The information can be generated quickly without need to wait development team to develop the query to generate report based on executive management requirement. With data

Fig. 9. Dashboard Analysis

warehouse developed for XYZ company, executive management can see how much the sales they achieve based on time (month, year), how is the performance of each payment method or payment gateway and which destination or departure airport is favorite based on customers behavior.

Recommendation for next research is to integrate the hotel booking system with current data warehouse developed. This might help executive management to get more insight on each business area they have. System also can be enhanced in the future to do prediction analysis based on customers behavior to help executive management on marketing field.

REFERENCES

[1] T. K. Das and A. Mohapatro, "A Study on Big Data Integration with Data Warehouse," International Journal of Computer Trends and Technology, vol. 9, pp. 188-192, 2014.

[2] A. M. Sundjaja, "Implementation of Business Intelligence on Banking, Retail and Educational Industry," Communication and Information Technology (COMMIT), vol. 7, no. 2, pp. 65-70, 2013.

[3] A. S. Girsang and A. Purwanto, "Controlling system for stock raw material for production planning and inventory control in a pharmacy company," International Review of Mechanical Engineering (IREME), vol. 11, no. 11, pp. 855-861, 2017.

[4] D. Martins, C. Ramos, J. Rodrigues, P. Cardoso, R. Lam and F. Serra, "Challenges in Building a Big Data Warehouse Applied to the Hotel Business Intelligence," in 6th Int. Conf. on Applied Informatics and Computing Theory (AICT'15), in Recent Research in Applied Informatics, Salerno, Italy, 2015.

[5] C. Huda, R. P. Sari, M. H. Haekal and T. Agustaria, "The Development of the Application of Data Warehouse at PT JKL," Communication and Information Technology (COMMIT), vol. 6, no. 1, pp. 8-19, 2012.

[6] M. N and S, "A Survey of ETL Tools," International Journal of Computer Techniques, vol. 2, no. 5, pp. 20-27, 2015

[7] P. G. H. a. R. P., "ETL Data conversion: Extraction, transformation and loading Data conversion," International Journal Of Engineering And Computer Science, pp. 22545-22550, 2017.

[8] M. R and K. P., "A Comparative Review Of Data Warehousing ETL Tools With New Trends And Industry Insight," in IEEE 7th International Advance Computing Conference, Hyderabad, 2017.

[9] C. W. Prakoso and A. S. Girsang, "Data Warehouse Development for Customer WIFI," International Journal on Communications Antenna and Propagation, vol. 7, no. 2, pp. 114-124, 2017.

[10] K. Ralph and R. Margy, The data warehouse toolkit: the complete guide to dimensional modeling., John Wiley & Sons, 2011.

[11] K. F. Tagirova and A. R. Ramazanov, "Data warehouse for intelligent oil wells control systems," in IEEE, St. Petersburg, Russia, 2017.

[12] T. Abdulaziz, I. Moawad and W. M. Abu-Alam, "Decision Support System Utilizing Data Warehouse Technique for the Tourism Sector in Egypt," in The 7th International Conference on Information Technology, Amman, Jordan, 2015.

[13] I. Moalla, A. Nabli, L. Bouzguenda and M. Hammami, "Data warehouse design from social media for opinion analysis: The case of Facebook and Twitter," in 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), Agadir, Morocco, 2016.

[14] C. Kraus and R. Valverde, "A data warehouse design for the detection of fraud in the supply chain by using the Benford's Law," American Journal of Applied Sciences, vol. 11, no. 9, pp. 1507-1518, 2014.

[15] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy and B. Becker, The Data Warehouse Lifecycle Toolkit, New Jersey: Wiley, 2007.

Check 2018, Alfredo.pdf

ORIGINALITY REPORT

23%	6%	18%	16%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Brawijaya Student Paper	4%
2	Abba Suganda Girsang, Danang Satya, Sani Muhamad Isa, Salman Al Fariz, Bambang Susilo, Dudi Ramdani, MaxLian. "Decision support system using data warehouse for hotel reservation system", 2017 International Conference on Sustainable Information Engineering and Technology (SIET), 2017 Publication	3%
3	www.apcorise.com Internet Source	3%
4	Ivan Alexander, Rian Rassetiadi, Samuel Garcia, Abba Suganda Girsang, Sani Muhamad Isa. "Business Solution for Choosing Products Using Data Warehouse in Payment Solution", 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), 2018 Publication	2%

5	Submitted to Stevens Institute of Technology Student Paper	2%
6	Submitted to Laureate Higher Education Group Student Paper	2%
7	K. F. Tagirova, A. R. Ramazanov. "Data warehouse for intelligent oil wells control systems", 2017 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), 2017 Publication	1%
8	Fakhri Nurullah, Gunawan Wang, Emil Robert Kaburuan, Ahmad Nurul Fajar. "The Collaboration of DevOps Automation and SOA to Accelerate Software Development Culture", 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), 2018 Publication	1%
9	paper Student Paper	1%
10	Elzbieta Malinowski. "Introduction to Databases and Data Warehouses", Advanced Data Warehouse Design, 2008 Publication	1%
11	en.wikipedia.org Internet Source	1%

12	fenix.tecnico.ulisboa.pt Internet Source	1%
13	Submitted to Northcentral Student Paper	1%
14	Submitted to Webster University Student Paper	1%
15	Abba Suganda Girsang, Evans Andita, Sani Muhamad Isa, Arie Purnama, Aditya, Ferico Samuel. "Business intelligence for physical examination platform service reporting system", 2017 International Conference on Sustainable Information Engineering and Technology (SIET), 2017 Publication	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		